Article

Public deliberations in the Council of the European Union: Introducing and validating DICEU

Christopher Wratil [iD]

Minda de Gunzburg Center for European Studies, Harvard University, Cambridge, MA, USA; Cologne Center for Comparative Politics, University of Cologne, Cologne, Germany

Sara B Hobolt

Department of Government, London School of Economics and Political Science, London, UK

Abstract

The Council of the European Union is the European Union's most powerful legislative body. Yet, we still have limited information about Council politics because of the lack of suitable data. This paper validates a new approach to studying Council politics entitled *DICEU – Debates in the Council of the European Union*. This approach is the first to leverage the public videos of Council deliberations as a data source. We demonstrate the face, convergent, and predictive validity of DICEU data. Governments' ideal points scaled from these videos yield meaningful and well-known conflict dimensions. Moreover, governments' positions during Council negotiations correlate highly with expert assessments and predict subsequent votes on legislative acts. We conclude that DICEU data provide a promising new approach to studying Council politics and multilevel governance.

Corresponding author:
Christopher Wratil, Minda de Gunzburg Center for European Studies, Harvard University, Adolphus Busch Hall, 27 Kirkland Street, Cambridge, MA 02138, USA.
Email: cwratil@fas.harvard.edu

The Council of the European Union (EU) is the EU's primary legislative body, where national ministers discuss, negotiate, and vote on legislative proposals. Most research on politics in the Council is based on either expert interview data or voting records. These data sources have generated key insights into intergovernmental politics in the Council, exploring what factors drive governments' position-taking and voting (e.g. Bailer et al., 2015; Hagemann et al., 2017; Hagemann and Høyland, 2008; Thomson, 2011) and the nature of alignments and latent conflict dimensions in Council politics (e.g. Mattila, 2009; Plechanovova, 2011; Thomson, 2009; Thomson et al., 2004; Zimmer et al., 2005). They have also been used to test different bargaining models and estimate the distribution of power among member states and across EU institutions (e.g. Bailer, 2004; Costello and Thomson, 2013; Golub, 2012; Rasmussen and Reh, 2013; Thomson, 2008, 2011). However, existing data sources have key limitations when it comes to studying legislative politics in the Council. Expert data may be influenced by interviewee bias and are hard to replicate for other researchers (e.g. Bueno de Mesquita, 2004). Voting data, in turn, tell us little about the preceding negotiation process and lack variation, since 98% of all votes are cast in favor of the proposal (Hagemann et al., 2017). In addition, none of the available datasets on Council politics captures the evolution of negotiations on legislative files over time.

In this article, we propose a novel approach to studying the Council, which uses video recordings of public deliberations. Since the entry into force of the Lisbon Treaty, the Council must deliberate in public when negotiating legislative files and discussing strategic questions[1] and all debates are video-streamed online. We present the first study that rigorously collects, codes, and validates these public deliberations as a data source to study the Council. We label our data gathering approach *DICEU* – "*Debates in the Council of the European Union*". The approach produces textual data of all speeches by national ministers and other representatives in the Council as well as human codings of intergovernmental conflicts.

We validate the DICEU approach in three steps. First, we demonstrate its face validity with a pilot dataset covering Council meetings in the Economic and Financial Affairs Council (ECOFIN) configuration between 2010 and 2015. Scaling governments' ideal points with an item-response theory (IRT) model as well as different quantitative text analysis (QTA) models yields meaningful latent conflict dimensions that are in line with prior knowledge about politics in the Council. High correlations between the ideal points scaled from human coding and those from text highlight the potential of QTA models for future analyses of Council politics. Moreover, we demonstrate DICEU's potential for generating further discoveries by tracing the dynamic development of conflict in the

Council over time as well as measuring new quantities of interest (such as ministers' sentiment or issue attention). Second, we assess convergent validity with expert data from the Council, using the "Decision-making in the European Union" (DEU) dataset as a benchmark (Thomson et al., 2006, 2012). We find that DEU and DICEU positions on the exact same issues correlate highly, which suggests that the public deliberations allow us to capture governments' negotiation positions. Finally, we investigate predictive validity of our position measures on Council votes. We find that governments' DICEU positions are a significant predictor of their voting behavior in the Council, even when controlling for factors that are known to influence voting.

In conclusion, we argue that the DICEU approach offers a rich data source on Council deliberations that can be used to study many aspects of politics in the EU and its member states. To facilitate further research, we make our entire DICEU pilot dataset on ECOFIN (2010–2015), as well as our DICEU codebook and related documentation, publicly available.

Data on Council politics

The vast majority of empirical research on politics in the Council relies either on expert interviews (e.g. Kleine, 2013; Thomson et al., 2006, 2012; Wasserfallen et al., 2019) or on official voting records (e.g. Hagemann, 2007; Hagemann and Høyland, 2008; Mattila, 2004, 2009) as the main data source. In addition, some work has analyzed official documents from the Council (e.g. Häge, 2014; Obholzer, 2014; Sullivan and Selck, 2007) and the policy statements that governments can submit with their voting record (Hagemann et al., 2019). While these data have produced important insights, they also have limitations.

The drawbacks of relying on expert data on the Council are well known: first, expert interviews must be conducted in close temporal proximity to negotiations in the Council to reduce the potential problem of post-diction bias (Bueno de Mesquita, 2004). The more distant the memory of negotiations, the more likely are experts to use cues, such as the negotiation outcome or their personal political convictions, when reconstructing actor positions. This likely introduces bias. Second, the social scientific standard of replicability is ill-defined in the case of expert interviews. While analyses of voting data, in contrast, are easily replicable, they only reflect the very last phase of Council politics, when legislation is passed, and cannot tell us about the actual proceedings of negotiations. Moreover, since a strong consensus norm in Council voting prevails (Heisenberg, 2005; Novak, 2013), with less than 1.5% of votes cast in opposition to legislative proposals in the 1999–2011 time period, variation in voting data is very limited (Hagemann et al., 2017). These consensual votes thus give relatively limited information about government positions during the negotiations. A more recent innovation in the literature on Council politics is large-N data created on the basis of official Council documents; for instance, draft minutes of Council meetings as well as presidency notes on the state of negotiations (see Häge, 2014; Obholzer, 2014).

One concern with this approach is that we know little about how the national presidencies may differ in the degree to which they make Council documents public as well as the amount of information they include in these documents (see also Hillebrandt et al., 2014).

Finally, existing data sources also do not allow us to study the dynamics of the negotiation process, as they are usually snapshots of one time point. Even the most extensive expert project on the Council, the DEU data (Thomson et al., 2006, 2012), does not contain sufficient data to estimate how government positions on key conflict dimensions (e.g. left–right, pro-anti regulation, pro-anti integration) develop over time. As a result, we still lack dynamic time-series-cross-section data in Council research. This prevents us from investigating research questions on the dynamics of Council negotiations, and it also makes the causal identification of effects more difficult, since many causal inference methods rely on temporal ordering.

To address some of these limitations, we present a new complementary approach to studying Council politics. We argue that videos of public deliberations of the Council provide a promising, and hitherto largely unexplored, data source for understanding the Council. While the Council has historically been a largely opaque institution of EU-level policy-making detached from the public eye, it has gradually become more transparent (see e.g. Hillebrandt et al., 2014). In June 2006, the European Council decided to make a large proportion of its legislative deliberations in the Council open to the public (European Council, 2006). From late 2006, webcasts of the open deliberations were made available on the Council's website in 20 different languages. Since the adoption of the Lisbon Treaty in December 2009, the Council configurations have to meet in public when they deliberate or vote on draft legislative acts, or discuss presidency work programs and other questions of strategic relevance. All these meetings must be broadcast on the Council's video streaming website (video.consilium.europa.eu).

While Council deliberations are now video-recorded, this increased transparency has not turned the Council into a classical legislature where representatives address a public audience in plenary debates. Instead, ministers' speeches during Council deliberations are much more technical and forthright than most parliamentary speeches (see the Online appendix for examples). Ministers discuss the details of policy-making and they openly advise colleagues on how to deal with domestic voters. This suggests the videos show actual negotiations, or at least part thereof.[2]

Videos of public deliberations overcome key limitations of existing data sources. They are not affected by post-diction bias, such as interviews may be, and they can be analyzed at any point in time (if archived). They also provide the most direct measurement of actor positions, free of any information filters applied by experts or officials. Compared to Council documents, such as draft minutes of Council meetings, they provide a full account of all public deliberations rather than short summaries. Analyses using video data are also fully replicable, as researchers have access to the exact same source material. Importantly, this is also a very rich data

source, as the Council produces around 100 hours of video material per calendar year, which allows researchers to observe the dynamics of policy-making in the Council across a large number of policy issues.

The DICEU approach

To study the Council using videos of public deliberations, we have developed the DICEU data gathering approach based on two components: textual data of ministers'/representatives' speeches and human codings of these speeches. The data extraction process firstly involves the transcription of the video footage of public deliberations in English[3] (since meetings are not transcribed) and secondly, the identification and coding of debates according to a comprehensive codebook, which contains variables on the debate topic, characteristics of the speaker, the language of the original/non-translated speech, and the general approval a speaker expresses during a debate. Full details are in the Online appendix.

The most important information contained in DICEU data for substantive purposes relates to the speaker's general approval of the proposal(s) being negotiated. We therefore introduce the *general approval* variable in some depth here. To generate this variable, coders identify a major dimension of contestation within each debate. For instance, this may be "Approval of the Commission's proposal on X". Due to rather uniform practices used by Council presidencies to structure and moderate debates, this dimension is almost without exception about (dis)approval of either (a) a legislative proposal presented by the European Commission; (b) the Council state of play on a proposal (e.g. a presidency compromise proposal, negotiations at lower bureaucratic levels); or (c) the state of negotiations with the European Parliament (EP) (e.g. a presidency's mandate for negotiations with the EP). In fact, many presidencies close their entry statement by asking whether delegations "approve" of a proposal or compromise proposal. Coders use a five-point Likert item to assess the degree of approval of each actor speaking in the debate:

1. The speaker expresses full approval.
2. The speaker expresses more approval than disapproval.
3. The speaker expresses a balance of approval and disapproval.
4. The speaker expresses more disapproval than approval.
5. The speaker expresses full disapproval.
6. Degree of approval cannot be assessed (applies to very short or irrelevant interventions).

The resulting dataset (excluding descriptive variables, such as the position of the speaker) can be thought of as a $D \times A$ matrix, in which D denotes debates and A denotes the actors/speakers. Each element of the matrix is a coding of the actor on the general approval item. We assessed the inter-coder reliability of our human codings using Krippendorff's alpha. For the general approval variable that is

measured on the same scale in each debate, alpha is 0.82, based on a random sample of double-coded debates (see the Online appendix). In the following, we assume that national delegations are unitary actors. Hence, we analytically pool all speakers giving speeches for the same national delegation (e.g. the French permanent representative and the French minister).[4] Dimension A therefore has 28 columns for the national delegations plus one column for the European Commission that usually participates in debates (e.g. introduces a new legislative proposal) and one column for "other" actors (e.g. in ECOFIN usually the European Central Bank (ECB) that contributes to some debates). The resulting text corpus containing the speech transcriptions is likewise a $D \times A$ matrix, in which each element is all text produced by an actor in one debate, i.e. we simply pool all speech interventions if an actor speaks more than once during a debate.

To validate the DICEU approach, we have applied it to all public deliberations held in the ECOFIN Council configuration between 2010 and 2015. From mid-2011 onwards, all videos of public deliberations are available on the Council's video streaming website. In addition, we were able to obtain the videos of several meetings held in 2010 and the first half of 2011 from the Council Secretariat. We focus on the ECOFIN configuration in this time frame for two reasons: first, economic and financial affairs is a policy area in which we have much prior knowledge of governments' preferences and available criteria to cross-validate our data with. For instance, we expect to see differences in the positions of net contributors and net beneficiaries of the EU budget (e.g. Bailer et al., 2015). Likewise, we anticipate a divide between creditors and debtors during the Eurozone crisis. Second, during this period, ECOFIN dealt with a high number of very salient and important legislative files and was arguably the main locus of EU secondary law-making activity. Prominent legislative packages include the banking union, the "two-pack" and "six-pack", and several issues on taxation (such as the financial transaction tax) as well as the always salient annual budget negotiations. In addition to our ECOFIN data, we also collected further data on a selection of files negotiated between 2006 and 2008 for further validation purposes with expert data, detailed below.

In total, our DICEU dataset from ECOFIN comprises of 1195 debate participations by national delegations, the Commission and other actors (e.g. the ECB) from 89 debates.[5] Debates lasted 42 minutes on average, but vary considerably in length with a standard deviation of almost 37 minutes. The median number of words spoken by a representative during a debate is 242 and the mean is 377. Across all debates, the British delegation spoke more than any other national delegation (about 29,000 words), but less than a third of the 91,000 words spoken by Commission representatives. The Online appendix contains further details on the data, including various descriptive statistics as well as a detailed analysis of the missing data patterns in DICEU data, which shows that the underlying missing data processes are likely "ignorable". In the subsequent sections, we evaluate the validity of the key measures included in DICEU data.

Face validity: Ideal points from human coding and text

We first assess *face validity*, i.e. the extent to which DICEU data appears to capture legislative politics in the Council accurately, by scaling governments' ideal points[6] from our ECOFIN debates between 2010 and 2015. To measure governments' ideal points, we use our human codings of approval as well as our debate transcriptions. We assume that the conflict space in the ECOFIN Council is unidimensional, but this assumption could be relaxed in future work. To scale governments' ideal points based on the general approval variable, we use a Bayesian mixed factor analysis model that is equivalent to a standard two-parameter IRT model with a probit link, when the items are ordinal as in our case (Quinn, 2004). The model assumes that the observed responses x_{ij}, where i denotes governments/actors and j denotes items, are determined by a matrix of latent variables x_{ij}^* as well as cut-points γ (Quinn, 2004: 339). In our case, this simplifies to

$$x_{ij} = c \tag{1}$$

$$x_{ij}^* \in [\gamma_{j(c-1)}, \gamma_{jc}] \tag{2}$$

where c indexes the categories in our general approval items and takes values from 1 (full approval) to 5 (full disapproval). Equation (2) imposes a strict ordinal relationship between the latent responses and the categories of our items. The factor analytic model for the ideal points of the governments derives as

$$X_i^* = \Lambda\theta_i + \varepsilon_i \tag{3}$$

where X_i^* is the vector of government i's latent responses, Λ is a matrix of factor loadings of the items on the estimated factors, θ_i is the vector of governments' factor scores, and ε_i is the error term. Setting the first element of θ_{i1} to 1 turns the first element of Λ_{j1} into the item's difficulty parameter, Λ_{j2} into the item's discrimination parameter, and θ_{i2} into the government's factor score on the first dimension, i.e. her ideal point (Armstrong et al. 2014: 296). The model is fitted via a Markov Chain Monte Carlo (MCMC) algorithm (further details are in the Online appendix).

We also analyze the transcriptions of the debates without any form of human coding using two closely related unsupervised text scaling models as well as sentiment analysis. First, we employ the standard Wordfish Poisson text scaling model (Slapin and Proksch, 2008), which has already been used for applications in EU politics (e.g. Franchino and Mariotto, 2012). We simply pool all interventions by a government across all debates in one document. Second, we use the Wordshoal model recently proposed by Lauderdale and Herzog (2016), which is a two-stage procedure relying on Wordfish estimates for each debate that are subsequently used as data in a Bayesian factor analysis. The Wordfish estimates of a government

in the debates it participated in, ψ_{ij}, where j denotes the debate, are used as data in the second stage and assumed to be a linear function of the government's ideal point, θ_i (see Lauderdale and Herzog, 2016: 378)

$$\psi_{ij} \sim N(\alpha_j + \beta_j \theta_i, \tau_i) \tag{4}$$

where α_j is a debate fixed effect and β_j a debate loading (or debate marginal effect) that indicates how strongly a debate is associated with the common latent dimension. The Wordshoal model is fitted using an MCMC algorithm for the second stage. Third, we also use the Lexicoder sentiment dictionary (Young and Soroka, 2012) to calculate a standard measure of "tone" for each government. Tone simply is the difference between positive (+ negated negative) and negative (+ negated positive) words divided by all words.

In total, we estimate four different types of ideal points for each government: *IRT approval* is an estimate from human coding of a government's approval of the state of negotiations (general approval variable), *Wordfish* is the Poisson text scaling estimate, *Wordshoal* is the two-stage estimate with Wordfish for each debate and a Bayesian factor analysis on top, and *tone* is the Lexicoder sentiment. Table 1 displays correlations between the four types of ideal points for our 28 governments.

The correlations are strikingly high between 0.73 and 0.95. It is particularly noteworthy that we find a high correlation of 0.81 between Wordfish and IRT approval. This suggests that models based on Poisson text scaling primarily capture to what extent governments approve of the state of play of negotiations in the Council. Hence, if researchers are only interested in a general conflict dimension about the approval of EU legislation, they may reliably use Wordfish (or Wordshoal) in future projects.

Next, we investigate more closely the actor alignments on the main conflict dimension in public ECOFIN deliberations by plotting governments' estimated ideal points from the IRT approval model and the Wordfish model in Figure 1 (a) and (b). Figure 1(c) shows a strong correlation between the two types of ideal points. The figure also reveals a clear north–west versus south–east cleavage with the United Kingdom, Sweden, the Netherlands, or Denmark among the least approving of ECOFIN politics and countries such as Greece, Cyprus, Estonia, or Croatia among the most approving. Figure 1(d) confirms the findings from the approval IRT by demonstrating that the member states in the north–west are also

Table 1. Correlations between four types of ideal points.

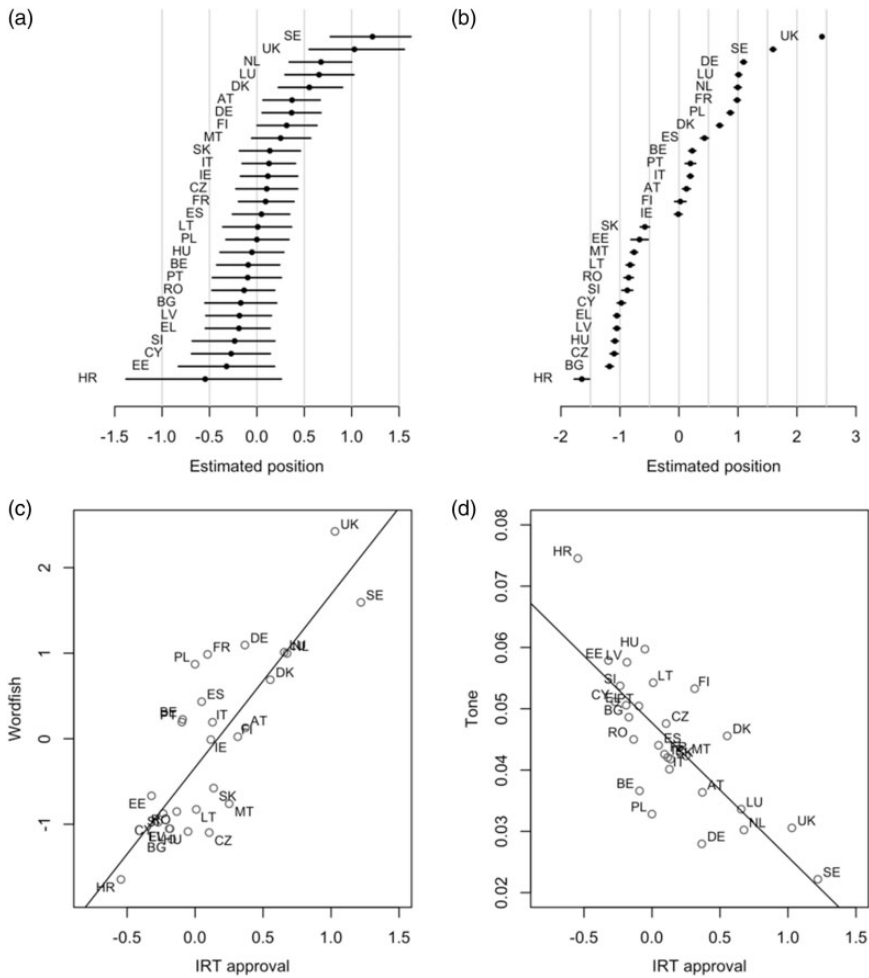|	IRT approval	Wordfish	Wordshoal	Tone
IRT approval	1	0.81	0.77	−0.78
Wordfish	0.81	1	0.95	−0.81
Wordshoal	0.77	0.95	1	−0.73
Tone	−0.78	−0.81	−0.73	1

Figure 1. Comparison of ideal points from different scaling methods: (a) Ideal points from IRT approval, (b) Ideal points from Wordfish, (c) IRT approval and Wordfish, and (d) IRT approval and tone. Note: 95% credible/confidence intervals as vertical lines in panels (a) and (b); linear regression line in panels (c) and (d). AT: Austria; BE: Belgium; BG: Bulgaria; HR: Croatia; CY: Cyprus; CZ: The Czech Republic; DK: Denmark; EE: Estonia; FI: Finland; FR: France; DE: Germany; EL: Greece; HU: Hungary; IE: Ireland; IT: Italy; LV: Latvia; LT: Lithuania; LU: Luxembourg; MT: Malta; NL: The Netherlands; PL: Poland; PT: Portugal; RO: Romania; SI: Slovenia; SK: Slovakia; ES: Spain; SE: Sweden; UK: The United Kingdom. Since Croatia joined the EU in 2013, the measurement of its position is less reliable; the country is also placed less extreme in the Wordshoal model (not shown).

speaking in a more negative tone (in line with their disapproval of Council politics) than those from the south–east. These actor alignments replicate prominent findings in the literature on north–south, north–south–east, and east–west divides between governments in the EU (Mattila, 2009; Plechanovova, 2011; Thomson, 2011; Veen, 2011).

They also provide new insights into Council politics, as they suggest that in the context of ECOFIN negotiations, Western and Northern member states were the least approving of and most negative about legislative politics. In the Online appendix, we analyze the substantive meaning of the north–west versus south–east conflict dimension present in our sample of ECOFIN debates. This reveals that, broadly speaking, the rich member states in the north–west preferred smaller EU budgets and emphasized economic competitiveness, while the poorer states in the south–east preferred larger budgets and emphasized economic regulation.

An additional benefit of the DICEU approach is that it enables us to investigate the dynamics of legislative politics in the Council over time. As an example, we can track the relevance of the latent conflict dimension as scaled from the Wordshoal model over time. This allows us to asses when the actor alignments in public deliberations in ECOFIN were particularly aligned with the main latent conflict dimension. Or less technically, when was the north–west versus south–east divide most clearly visible? To answer this question, Figure 2 plots the absolute Wordshoal debate loadings, β_j, over time with a simple locally estimated scatterplot smoothing (LOESS) regression line. This illustrates that the conflict between member states in the north–west versus south–east became more prevalent in ECOFIN debates between mid-2012 and mid-2013 and remained important.[7] In the Online appendix, we also plot the debate-specific ideal points of all national governments from our Wordshoal results, $\psi_{ij} * \beta_j$, to further illustrate DICEU data's potential to investigate dynamics.

In sum, our analyses demonstrate a high level of face validity of DICEU data. Moreover, DICEU data provide us with new research opportunities, for instance, to investigate the Council deliberations in a temporal dimension or measure actors' attention to different political issues, which we illustrate in the Online appendix using dictionary approaches as an example.

Convergent validity: Comparison with expert data

Next, we assess *convergent validity*, i.e. the convergence of information retrieved from Council deliberations with existing measures expected to capture similar constructs. Given the limited data available on government positions in the Council, DEU data represents the closest we get to a "gold standard" for measurements of actors' positions in the Council to be used for cross-validation. Following the DEU project's assumption that experts report the most important issues related to a proposal as well as actors' initial negotiation positions on these issues (that is, immediately after the introduction of a proposal), we expect that DICEU
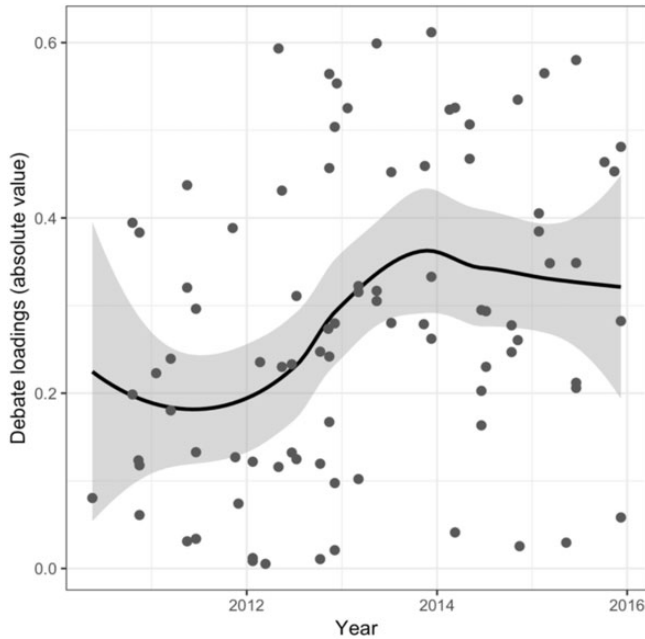
Figure 2. Absolute values of debate loadings from Wordshoal over time.
Note: Solid line is LOESS regression line with span = 0.75; dots are β_j.

positions should converge to information from DEU under the following assumptions:

1. Actors discuss the most important issues regarding a legislative proposal in public deliberations in the Council.
2. Actors reveal their negotiation positions in public deliberations.
3. Actors do not change their negotiation positions between the introduction of a proposal and its discussion in public deliberations.

We can check if (1) holds true by ascertaining how many issues from the DEU data are discussed in public deliberations. Moreover, assumption (3) should hold for Council deliberations that coincide with the introduction of a proposal or shortly thereafter. In case assumption (2) holds, we would thus expect to find high convergence between DEU and DICEU positions if the Council deliberations are held in close proximity to the introduction of the proposal. To ascertain this convergence, we identified a total of 21 legislative proposals covered in the DEU data that were placed as "B items" for first discussion on agendas of Council meetings after spring 2006, and could therefore, in principle, have been discussed in public. We were able to gather a dataset of re-digitized video footage from the Council's

archives covering public deliberations on 17 of these 21 proposals that were discussed in different Council configurations (not only ECOFIN).

This dataset provides us with a unique opportunity to cross-validate video-based with expert-based data. According to DEU experts, actors were divided on a total of 50 issues regarding these 17 proposals. Our coders identified a clear majority of 31 of these 50 DEU issues as discussion points in the videos of the public deliberations, and 12 of the DEU issues were discussed in more than one Council meeting. This demonstrates that there is some relationship between the main legislative issues identified by experts and the issues debated in public sessions of the Council. Furthermore, for the subset of issues covered by DEU and in the videos our coders placed actors on the DEU issue scales, based on their DICEU speeches. This allows us to directly compare information retrieved from experts and from videos on the same scale (i.e. the DEU issue scales).

Coders identified a total of 209 overlapping positions on 26 DEU issues, that is, in 209 cases, actors could be placed on the predefined DEU scales from 0 to 100 based on video footage _and_ DEU experts had reported their positions in interviews. The correlation of the DEU expert positions and the positions from the videos coded on the DEU scale is 0.60 with an average absolute distance between the positions of 23 points on the DEU scale. Hence, the positional information contained in DEU and DICEU is similar but clearly not identical. Yet differences between DICEU and DEU could stem from violations of assumption (3) as well as measurement error in both datasets instead of violations of assumption (2). Hence, we investigate whether convergence increases when public deliberations are held in close temporal proximity to the introduction of a proposal, and for particularly salient issues, where it should be easier for DEU experts and DICEU coders to accurately identify actors' positions.

Table 2 shows the results of a linear regression of the absolute distance between the DEU and the DICEU placements on our 209 overlapping positions on a variable indicating the time lag between the introduction of the proposal and the public deliberation (within six months, within 6–12 months, after 12 months) as well as the salience of an issue to the actor as reported by DEU experts. Two-way cluster-robust standard errors, accounting for the nesting of positions within actors and issues, are shown in parentheses. As expected, the results demonstrate that DEU and DICEU placements are much closer when DICEU codings are from public deliberations held within the first year of the introduction of the legislative proposal, and for salient issues.

In fact, if an issue was discussed within six months of the Commission's legislative proposal and the issue was highly salient to the actor (actor salience = 100), our model predicts DEU and DICEU measures to diverge by only 9.8 DEU scale points on average. Similarly, the correlation of positions reported by experts and those coded from videos rises to 0.78 for public deliberations within the first year after the Commission's proposal and with actor salience > 50. Hence, initial negotiation positions retrieved from expert interviews and positions coded from videos of public deliberations strongly converge when the issue is salient and the time lag

Table 2. Predicting the distance between DEU and DICEU positions.

	Estimates
Within 6–12 months	5.03
	(3.65)
After 12 months	13.41
	(5.98)*
Salience	−0.16
	(0.05)*
Constant	26.13
	(5.01)*
R^2	0.05
N	209

Note: Linear regression; cluster-robust standard errors by actors and issues in parentheses; *$p < 0.05$.

between the introduction of the proposal and the public deliberation is short. The DICEU approach can therefore provide similar information to the DEU approach about initial negotiation positions.

Predictive validity: Forecasting opposition votes

In addition to face and convergent validity, we also assess *predictive validity*, i.e. the question of whether variation in DICEU measures can forecast events; specifically, governments' voting behavior in the Council. This allows us to assess whether governments practice what they preach: are governments' statements in public deliberations an indication of their subsequent voting behavior on these acts in the Council?

For this purpose, we collected the voting sheets for a total of 95 legislative acts that were discussed in our 89 debates.[8] Out of these, six acts had not been decided on at the time of writing this article, five acts the Commission had withdrawn before voting, one act was non-legislative and not voted on, and for three acts we were unable to obtain voting results from the Council's register. Of the remaining 80 acts we observed 2194 votes by governments with 53 "No" and 30 "Abstain" votes. In line with common practice in the field, we pool no votes and abstentions as "opposition votes" (e.g. Bailer et al., 2015; Hagemann et al., 2017). Note that with 3.8% opposition votes, voting on ECOFIN files in this period is clearly more contentious than voting has been historically across Council configurations (averaging around 1–2% opposition votes). To assess the predictive validity of DICEU data for opposition votes, we use the DICEU general approval variable introduced above, which was coded at least from one debate for 1745 of our 2194 votes. Recall that this variable indicates the tendency to approve
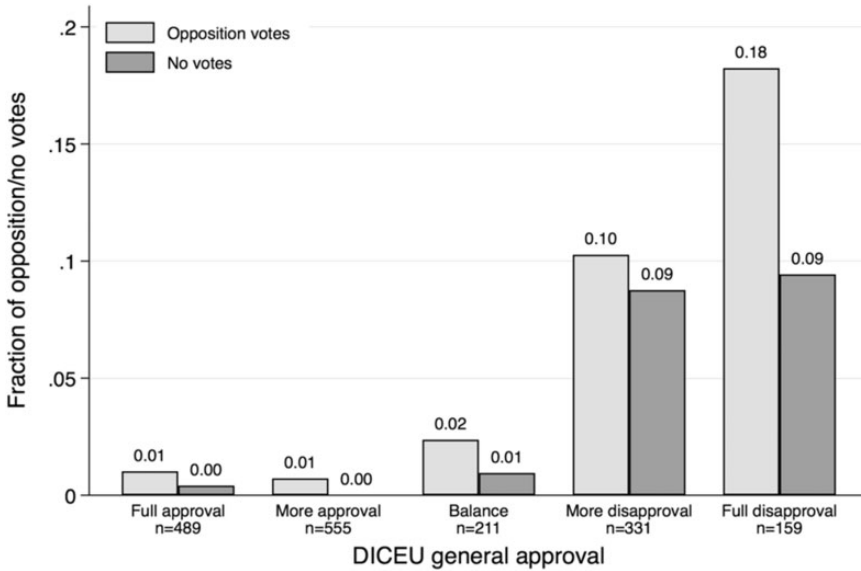
Figure 3. DICEU general approval and opposition/no votes.

of the state of negotiations. In Figure 3, we show the association between opposition/no votes and the general approval variable.[9]

This descriptive overview reveals a strong association between governments' approval of an act in the public deliberations and their later voting behavior on this act. Goodman and Kruskal's gamma is 0.72 for the association between opposition votes and approval and 0.78 for the association between no votes and approval. If governments expressed more approval than disapproval or even full approval, 1% or less of the votes were opposition or no votes. However, if governments on balance expressed more disapproval or even full disapproval, the fraction of subsequent no votes rose to 9% and that of opposition votes to 18%. This illustrates the basic predictive validity of DICEU measures for governments' voting behavior.

But do DICEU data simply capture factors that we know from the literature predict opposition votes, or do they contain novel information that contributes to predicting opposition votes while controlling for other factors? To answer this question, we run logistic regression models with the binary indicator of opposition votes as dependent variable, the *DICEU disapproval* measure (running from 1 "full approval" to 5 "full disapproval") as main independent variable, and controls from a standard Council voting model as in Hagemann et al. (2017). Following Hagemann et al. (2017), we include a measure of *Public opinion* on European integration from the Eurobarometer series, specifically the "EU image" question, to detect signaling to domestic audiences (see also Wratil, 2018b). To capture the

Table 3. Logit models predicting opposition votes in the Council.

	Model 1	Model 2
DICEU disapproval	0.54	0.38
	(0.13)*	(0.18)*
Public opinion	−2.74	−4.84
	(0.66)*	(1.63)*
Government position EU integration	−0.17	0.10
	(0.08)*	(0.19)
Government position left–right	0.02	−0.01
	(0.01)	(0.01)
Receipts from EU budget	−0.25	0.15
	(0.15)	(0.63)
Unemployment rate	−0.13	−0.42
	(0.05)*	(0.17)*
Inflation rate	0.10	0.25
	(0.10)	(0.13)
Budget acts	0.88	1.18
	(0.40)*	(0.44)*
Constant	4.03	12.05
	(2.19)	(6.15)
Number of acts	80	80
Number of countries	28	18
Country fixed effects	No	Yes
N	2194	1436

Note: Logistic regressions; standard errors in parentheses; *$p < 0.05$.

"party politics" thesis, we use measures of *Government parties' positions on EU integration and left–right* at the time of the last election from the Comparative Manifesto Project's (CMP) coding of election manifestos (Volkens et al., 2018; Wratil, 2018a). Further, we account for economic interest explanations with measures of the *Receipts from the EU budget* (in % of national gross domestic product) as well as the national *Unemployment* and *Inflation rates* (Bailer et al., 2015). Last, we use a dummy variable for votes on *Budget acts* (e.g. draft amending annual budgets), since opposition votes on budget acts are much more frequent than on other legislative files (74 out of the 83 opposition votes occurred on acts concerning the EU budget). We impute missing values using chained equations with predictive mean matching (see the Online appendix for full details on variables and imputation).

In Table 3, we report the results from a baseline model (Model 1) and a model adding country fixed effects (Model 2), which reduces our sample to just 18 countries that casted at least one opposition vote. The results demonstrate that the DICEU approval variable is a strong predictor of opposition votes even when controlling for other factors from the literature. The more disapproval

governments show in public deliberations, the more likely they are to cast an opposition vote. Hence, DICEU positions do not only capture the effect of other variables but provide us with new information to predict Council voting. In addition, we find support for the "signaling" explanation of Council votes, since the coefficient on public opinion is a statistically significant predictor, indicating that governments facing more Eurosceptic publics cast more opposition votes. In contrast, we find little support for the "party politics" explanation, except for one significant coefficient on the government parties' position on EU integration.

In sum, our results demonstrate that DICEU positions have strong predictive validity with regard to Council voting and a significant part of the information they provide is novel and cannot be substituted by common other predictors of Council voting from the literature.

Conclusion

In this paper, we have introduced and validated a new data source and approach for studying politics in the Council of the European Union. The DICEU approach extracts and codes textual data from videos of public deliberations in the Council, a data source not used in any publication in political science so far. We have addressed the face, convergent, and predictive validity of data gathered with this approach.

Studying the Council from videos of its public deliberations has obvious advantages. Most importantly, it is the most direct way of measuring government positions during actual negotiations, since it does not rely on the recollection or interpretation by third parties (e.g. experts, presidencies). This approach can also be applied by researchers across the world at any time and the analyses are fully replicable.[10] Video data also bridge the gap between the well-established expert data projects (such as DEU) that measure initial negotiation positions before bargaining takes place and voting data that reveal government positions after negotiations are concluded. Video data essentially open up the black box of Council negotiations, although – as with any other forum for political negotiation – we can never be certain what is happening informally behind closed doors.

Our pilot dataset on public deliberations in ECOFIN can be easily extended in the future. The finding that QTA estimates (e.g. from Wordfish and Wordshoal) correlate highly with estimates from models based on human coding suggests that collecting transcriptions of debates without extensive and costly human coding may be sufficient for many research projects. Proksch et al. (2019) have recently validated the use of automatic speech recognition systems to transcribe video and audio material for QTA. Based on these findings, future projects can analyze the abundance of available video data from further Council configurations and other time frames by simply feeding it into an automatic speech recognition system (e.g. Google's Speech application programming interface or YouTube) and analyzing retrieved transcriptions through QTA models. Our results indicate that Poisson text scaling models will retrieve a conflict dimension that largely reflects

the approval of EU legislation by governments. The interpretation of quantities of interest from any other text models should be validated by researchers (e.g. on the basis of hand-coded sub-samples).

DICEU data will enable researchers to address new research questions about Council politics and multilevel governance. Crucially, many questions about dynamics can now be addressed, as changes in actors' positions can be analyzed. Moreover, in the ideal point framework, changes in positions over time can be identified by disentangling debate- or period-specific effects from genuine movements on the latent dimension (see Martin and Quinn, 2002). Dynamic ideal point models are either available already (e.g. Martin and Quinn, 2002; Schnakenberg and Fariss, 2013) or can be easily developed from their existing static versions (Lauderdale and Herzog, 2016: 392). We can thus use DICEU data to address questions such as: when and why are governments changing positions in the Council and what are the effects on the bargaining process and outcomes?

Beyond the dynamic perspective, the linguistic characteristics of text from video data provide new opportunities for innovative projects. These are not limited to the application of standard scaling models like Wordfish and Wordshoal, but could also include dictionary-based approaches to sentiment, complexity, or technicality of speech, as well as topic models to measure issue attention or arguments (e.g. Grimmer, 2010; Roberts et al., 2014). Moreover, the growing methodological literature on the analysis of audio-visual features (e.g. Dietrich et al., 2018; Knox and Lucas, 2018; Schonhardt-Bailey, 2017), such as facial expressions or vocal pitch, suggests that DICEU data could be used in the future to systematically measure these novel aspects of political communication in the Council.

As any data source, videos of public deliberations have their limitations. One weakness of these data compared to expert data is that it will always be finite. Not all national delegations speak in all debates and on all topics. Hence, if researchers are interested in coding government positions on very specific issues, they may encounter missing data. The underlying missing data process is likely ignorable (see the Online appendix), and for many analyses that are prone to missing data, the use of multiple imputation techniques will suffice to obtain unbiased estimates. However, where analysis models are very sensitive to missing data, such as tests of competing bargaining models (e.g. Slapin, 2014; Thomson, 2011), this may present a limitation of the DICEU approach.

Our choice of data sources and collection strategies should be guided by the questions we ask. We believe that many key questions in European politics can be addressed by studying videos of public deliberations in the Council, and by making the DICEU dataset publicly available, we hope that researchers will be able to answer some of these.

Acknowledgments

Brooks, Sarah Ciaglia, Carolyn Konopka, Jason Krstic, Lukas Lehner, Ida Popovski, and Tim Rogers for their excellent research assistance and to Felix Bethke, Sven-Oliver Proksch, and Gerald Schneider for helpful comments on earlier versions of this article. The entire DICEU dataset used in this article is included in the replication files and more information on the data is available online at www.wratil.eu.

Funding

ORCID iD

Christopher Wratil [iD] http://orcid.org/0000-0002-7339-9628

Notes

1. Note that the Council can still deliberate in closed session over non-legislative issues such as implementation of legislation or preparations of international events (e.g. G7, G20).
2. As with any other legislature, we do not know what happens behind the scenes and to what extent public deliberations are representative of the broader process of negotiations (e.g. in COREPER, "Comité des représentants permanents", and the working groups).
3. When the speaker does not speak in English, we use the simultaneous translation into English provided by the Council's interpreters. Note that representatives can freely choose the language they use, since all speech is simultaneously translated into all official EU languages. Eighty-three percent of representatives choose to address the Council in English. Pooling speeches by English native speakers, non-native speakers using English as well as speakers translated into English best reflects the actual conduct of negotiations. De Vries et al. (2018) demonstrate that texts translated into other languages yield almost identical estimates in quantitative text models as the original texts. In the Online appendix, we also show that language choice made no difference to the human codings we obtained.
4. Note that according to our data about 70% of all speeches are given by national ministers. However, this information is primarily based on participant lists included in press releases and we found these to be inaccurate in many cases, without an obvious bias.
5. Note that we do not code any discussions with less than two national delegations (excluding the presidency) participating. Seventeen of such discussions occurred in our period of investigation. Hence, our data covers around 84% (89 out of 106) of all discussions in the Council, and the remaining 16% are by definition very short discussions with only a few participating actors.
6. In the ideal point models described below, we also estimate the positions of the European Commission and "other" speakers (usually the ECB). But we do not report their positions. It turns out that due to specificities of the human coding instructions, these actors have very different ideal points when scaled from human coding as opposed to speeches. Re-estimating all models without these actors yields substantively the same high correlations between governments' ideal points.

7. Obviously, this increased relevance can be due to changing actor alignments, or alternatively, to a changing political agenda, e.g. new legislative issues that more strongly mobilize the north–west versus south–east divide. We therefore hand-coded debates and separated them into five different issue areas. This reveals that the debate loadings showed upward trends over time in four out of these five areas, suggesting that the increased relevance of the latent conflict dimension is not due to certain issues reaching the political agenda.

8. Note that a single debate can deal with no (e.g. on the presidency work program), one, or several specific legislative acts (e.g. debate on banking union).

9. If more than one debate was held on an act and approval was coded for more than one debate, we take its coded value from the chronologically first debate on an act. In the Online appendix, we report results using the last debate on an act.

10. Video material is rarely used to systematically measure government positions in international negotiations, but one exception is McKibben's (2016) study of multilateral climate change negotiations using archived webcasts.

References

Armstrong DA, Bakker R, Carroll R, et al. (2014) *Analyzing Spatial Models of Choice and Judgment in R*. Boca Raton: CRC Press.

Bailer S (2004) Bargaining success in the European Union: The impact of exogenous and endogenous power resources. *European Union Politics* 5(1): 99–123.

Bailer S, Mattila M and Schneider G (2015) Money makes the EU go round: The objective foundations of conflict in the Council of Ministers. *Journal of Common Market Studies* 53(3): 437–456.

Bueno de Mesquita B (2004) Decision-making models, rigor and new puzzles. *European Union Politics* 5(1): 125–138.

Costello R and Thomson R (2013) The distribution of power among EU institutions: Who wins under codecision and why? *Journal of European Public Policy* 20(7): 1025–1039.

de Vries E, Schoonvelde M and Schumacher G (2018) No longer lost in translation. Evidence that Google translate works for comparative bag-of-words text applications. *Political Analysis* 26(4): 417–430.

Dietrich BJ, Enos RD and Sen M (2018) Emotional arousal predicts voting on the U.S. Supreme Court. *Political Analysis* 1–7.

European Council (2006) An overall policy on transparency. Available at: www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/misc/90112.pdf (accessed 13 March 2019).

Franchino F and Mariotto C (2012) Explaining negotiations in the conciliation committee. *European Union Politics* 14(3): 345–365.

Golub J (2012) How the European Union does not work: National bargaining success in the Council of Ministers. *Journal of European Public Policy* 19(9): 1294–1315.

Grimmer J (2010) A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis* 18(1): 1–35.

Häge FM (2014) Scaling policy preferences from negotiation positions. In: *7th Pan-European Conference on EU politics*, 5–7 June 2014, The Hague.

Hagemann S (2007) Applying ideal point estimation methods to the Council of Ministers. *European Union Politics* 8(2): 279–296.

Hagemann S, Bailer S and Herzog A (2019) Signals to their parliaments? Governments' use of votes and policy statements in the EU Council. *Journal of Common Market Studies*. Epub ahead of print 8 March 2019. DOI: https://onlinelibrary.wiley.com/doi/10.1111/jcms.12844.

Hagemann S, Hobolt SB and Wratil C (2017) Government responsiveness in the European Union: Evidence from Council voting. *Comparative Political Studies* 50(6): 850–876.

Hagemann S and Høyland B (2008) Parties in the Council? *Journal of European Public Policy* 15(8): 1205–1221.

Heisenberg D (2005) The institution of 'consensus' in the European Union: Formal versus informal decision-making in the Council. *European Journal of Political Research* 44(1): 65–90.

Hillebrandt MZ, Curtin D and Meijer A (2014) Transparency in the Council of Ministers of the EU: An institutional approach. *European Law Journal* 20(1): 1–20.

Kleine M (2013) *Informal Governance in the European Union. How Governments Make International Organizations Work*. Ithaca: Cornell University Press.

Knox D and Lucas C (2018) A dynamic model of speech for the social sciences. Working Paper.

Lauderdale BE and Herzog A (2016) Measuring political positions from legislative speech. *Political Analysis* 24(3): 374–394.

Martin AD and Quinn KM (2002) Dynamic ideal point estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999. *Political Analysis* 10(2): 134–153.

Mattila M (2004) Constested decisions: Empirical analysis of voting in the European Council of Ministers. *European Journal of Political Research* 43(1): 29–50.

Mattila M (2009) Roll call analysis of voting in the European Union Council of Ministers after the 2004 enlargement. *European Journal of Political Research* 48(6): 840–857.

McKibben HE (2016) To link or not to link? Agenda change in international bargaining. *British Journal of Political Science* 46(2): 371–393.

Novak S (2013) The silence of ministers: Consensus and blame avoidance in the Council of the European Union. *Journal of Common Market Studies* 51(6): 1091–1107.

Obholzer L (2014) *Bicameral Coalition Formation: Dynamics of Legislative Cooperation in the European Union*. London: London School of Economis and Political Science.

Plechanovova B (2011) The EU Council enlarged: North-south-east or core-periphery? *European Union Politics* 12(1): 87–106.

Proksch S-O, Wratil C and Wäckerle J (2019) Testing the validity of automatic speech recognition for political text analysis. *Political Analysis*. Epub ahead of print 19 February 2019. DOI: https://doi.org/10.1017/pan.2018.62.

Quinn KM (2004) Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* 12(4): 338–353.

Rasmussen A and Reh C (2013) The consequences of concluding codecision early: Trilogues and intra-institutional bargaining success. *Journal of European Public Policy* 20(7): 1006–1024.

Roberts ME, Stewart BM, Tingley D, et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science* 58(4): 1064–1082.

Schnakenberg KE and Fariss CJ (2013) Dynamic patterns of human rights practices. *Political Science Research and Methods* 2(1): 1–31.

Schonhardt-Bailey C (2017) Nonverbal contention and contempt in U.K. parliamentary oversight hearings on fiscal and monetary policy. *Politics and the Life Sciences* 36(1): 27–46.

Slapin JB (2014) Measurement, model testing, and legislative influence in the European Union. *European Union Politics* 15(1): 24–42.

Slapin JB and Proksch S-O (2008) A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* 52(3): 705–722.

Sullivan J and Selck TJ (2007) Political preferences, revealed positions and strategic votes: Explaining decision-making in the EU Council. *Journal of European Public Policy* 14(7): 1150–1161.

Thomson R (2008) The Council presidency in the European Union: Responsibility with power. *Journal of Common Market Studies* 46(3): 593–617.

Thomson R (2009) Actor alignments in the European Union before and after enlargement. *European Journal of Political Research* 48(6): 756–781.

Thomson R (2011) *Resolving Controversy in the European Union. Legislative Decision-Making before and after Enlargement*. Cambridge: Cambridge University Press.

Thomson R, Arregui J, Leuffen D, et al. (2012) A new dataset on decision-making in the European Union before and after the 2004 and 2007 enlargements (DEUII). *Journal of European Public Policy* 19(4): 604–622.

Thomson R, Boerefijn J and Stokman FN (2004) Actor alignments in European Union decision making. *European Journal of Political Research* 43(2): 237–261.

Thomson R, Stokman FN, Achen CH, et al. (eds) (2006) *The European Union Decides*. Cambridge: Cambridge University Press.

Veen T (2011) The dimensionality and nature of conflict in European Union politics: On the characteristics of intergovernmental decision-making. *European Union Politics* 12(1): 65–86.

Volkens A, Lehmann P, Matthieß T, et al. (2018) *The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR)*. Version 2018a. Berlin.

Wasserfallen F, Leuffen D, Kudrna Z, et al. (2019) Analysing European Union decision-making during the Eurozone crisis with new data. *European Union Politics* 20(1): 3–23.

Wratil C (2018a) *CMP Government Positions Data*. Version 1.1. Cambridge.

Wratil C (2018b) Modes of government responsiveness in the European Union: Evidence from Council negotiation positions. *European Union Politics* 19(1): 52–74.

Young L and Soroka S (2012) Affective news: The automated coding of sentiment in political texts. *Political Communication* 29(2): 205–231.

Zimmer C, Schneider G and Dobbins M (2005) The contested Council: Conflict dimensions of an intergovernmental EU institution. *Political Studies* 53(2): 403–422.